REVIEWS

Review
https://doi.org/10.36233/0372-9311-345

The Russian database of HIV antiretroviral drug resistance

Dmitry E. Kireev✉, Alina A. Kirichenko, Alexey E. Lopatukhin, Anastasia V. Shlykova, Nikolai Yu. Galkin, Evgenii V. Saveler, Maxim B. Glazov, Vadim V. Pokrovsky, Vasily G. Akimkin

Central Research Institute for Epidemiology, Moscow, Russia

Abstract

The development of sequencing technologies and bioinformatic analysis made it possible to conduct molecular and epidemiological studies, in which nucleotide sequences of the human immunodeficiency virus (HIV) are used as information added to the patient profile. From a practical perspective, studies of prevalence of HIV drug resistance (HIVDR) are of the highest significance. To promote such studies, different countries use databases that serve as repositories of genetic and epidemiological information. The Russian HIVDR database (https://hivresist.ru/) was created in 2009. Nevertheless, it was characterized by limited applicability for a long time. Since 2021, after the regulatory documents had been revised and updated, the entry of HIVDR research results into the Russian HIVDR database has been mandatory. Therefore, the priority attention has been given to upgrading the database and improving its functional capabilities. Different methods have been developed to enter clinical, epidemiological and genetic data. At the time of this study, the Russian database HIVDR contained 10,626 unique records about patients and 13,126 nucleotide sequences deposited by 10 institutions. The following functions have been provided for data analysis: quality control of the epidemiological and clinical information about a patient, quality control of nucleotide sequences, contamination check, subtyping, detection of DR mutations, identification of viral tropism and generation of standardized reports. The efforts toward further development of the Russian HIVDR database will be focused on designing tools for detection and analysis of molecular clusters, adaptation to routine application for epidemiological surveillance of HIV infection.

Keywords: *HIV, drug resistance, database, nucleotide sequence, quality control, subtyping, molecular clusters, molecular epidemiology*

Acknowledgements. The authors express their gratitude to Sergey Vladimirovich Shtrek, head of the laboratory at the Omsk Research Institute of Natural and Focal Infections of the Siberian Federal District Center for AIDS Prevention and Control; Natalia Nikolaevna Zaitseva, director, and Olga Yurievna Peksheva, head of the laboratory of molecular, genetic, and serological research methods at the Blokhina Nizhny Novgorod Research Institute of Epidemiology and Microbiology; Mikhail Valerievich Pitersky, head of the Ural District Center for AIDS Prevention and Control at the Ekaterinburg Research Institute of Viral Infections; Svetlana Vladimirovna Yaschenko, head of the laboratory of bacteriological, molecular, and genetic studies at the Krasnoyarsk Regional Center for AIDS Prevention and Control; Natalia Vladimirovna Filonyuk, head of the laboratory of clinical immunology at the Lipetsk Regional Center for Prevention and Control of AIDS and Infectious Diseases; Andrei Borisovich Shemshura, a doctor of clinical laboratory diagnostics at the Clinical Center for AIDS Prevention and Control at the Ministry of Health of the Krasnodar Territory; Dmitry Sergeyevich Kolpakov, head of the clinical and diagnostic laboratory at the Rostov Research Institute of Microbiology and Parasitology; Marina Ridovna Bobkova, head of the laboratory of leukemia viruses at the Gamaleya National Research Center of Epidemiology and Microbiology, Valeria Olegovna Kotova, head of the laboratory of epidemiology and prevention of viral hepatitis and AIDS at the Khabarovsk Research Institute of Epidemiology and Microbiology for depositing molecular, epidemiological and clinical information to the Russian HIVDR database.

Funding source. The study was funded through the inhouse grant of the Central Research Institute of Epidemiology of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor) for the fundamental research "Development of the Russian HIV-1 antiretroviral drug resistance database".

Conflict of interest. The authors declare no apparent or potential conflicts of interest related to the publication of this article.

For citation: Kireev D.E., Kirichenko A.A., Lopatukhin A.E., Shlykova A.V., Galkin N.Yu., Saveler E.V., Glazov M.B., Pokrovsky V.V., Akimkin V.G. The Russian database of HIV antiretroviral drug resistance. *Journal of microbiology, epidemiology and immunobiology.* 2023;100(2):219–227.
DOI: https://doi.org/10.36233/0372-9311-345 EDN: https://www.elibrary.ru/rrwanu

Научный обзор
https://doi.org/10.36233/0372-9311-345

Российская база данных устойчивости ВИЧ к антиретровирусным препаратам

Киреев Д.Е.✉, Кириченко А.А., Лопатухин А.Э., Шлыкова А.В., Галкин Н.Ю., Савельер Е.В., Глазов М.Б., Покровский В.В., Акимкин В.Г.

Центральный научно-исследовательский институт эпидемиологии Роспотребнадзора, Москва, Россия

Аннотация

Развитие технологий секвенирования и биоинформатического анализа дало возможность проведения молекулярно-эпидемиологических исследований, в которых нуклеотидные последовательности вируса иммунодефицита человека (ВИЧ) используются в качестве дополнительной характеристики пациента. При этом наиболее значимым с практической точки зрения направлением работ является изучение распространения лекарственной устойчивости (ЛУ) ВИЧ. В различных странах для организации таких исследований применяются базы данных, являющиеся хранилищами генетической и эпидемиологической информации. Российская база данных устойчивости ВИЧ к антиретровирусным препаратам (https://hivresist.ru/) была создана в 2009 г. Тем не менее длительное время её применение оставалось ограниченным. С 2021 г. после обновления нормативных документов внесение результатов исследований ЛУ ВИЧ в российскую базу данных устойчивости ВИЧ к антиретровирусным препаратам стало обязательным. В связи с этим были проведены работы по усовершенствованию базы данных и увеличению её функциональных возможностей. Были разработаны различные способы внесения клинико-эпидемиологических и генетических данных. На момент написания публикации российская база данных ЛУ ВИЧ содержала 10 626 уникальных записей о пациентах и 13 126 нуклеотидных последовательностей, загруженных 10 учреждениями. Для анализа данных были разработаны следующие функции: контроль качества эпидемиологической и клинической информации о пациенте, контроль качества нуклеотидных последовательностей, проверка на контаминацию, субтипирование, выявление мутаций ЛУ, определение вирусного тропизма и генерация стандартизированных отчётов. В планах по дальнейшему развитию российской базы данных ЛУ ВИЧ — разработка инструмента для выявления и анализа молекулярных кластеров и адаптация для рутинного использования в рамках эпидемиологического надзора за ВИЧ-инфекцией.

Ключевые слова: *ВИЧ, лекарственная устойчивость, база данных, нуклеотидная последовательность, контроль качества, субтипирование, молекулярные кластеры, молекулярная эпидемиология*

Благодарности. Авторы выражают благодарность Сергею Владимировичу Штреку, заведующему лабораторией Омского НИИ природно-очаговых инфекций Сибирского федерального окружного центра по профилактике и борьбе со СПИД; Наталье Николаевне Зайцевой, директору, и Ольге Юрьевне Пекшевой, заведующей лабораторией молекулярно-генетических и серологических методов исследования Нижегородского НИИ эпидемиологии и микробиологии им. академика И.Н. Блохиной; Михаилу Валерьевичу Питерскому, руководителю Уральского окружного центра по профилактике и борьбе со СПИД Екатеринбургского НИИ вирусных инфекций; Светлане Владимировне Ященко, заведующей лабораторией бактериологических и молекулярно-генетических исследований Красноярского краевого центра профилактики и борьбы со СПИД; Наталье Владимировне Филонюк, заведующей клинико-иммунологической лабораторией Липецкого областного центра по профилактике и борьбе со СПИД и инфекционными заболеваниями; Андрею Борисовичу Шемшуре, врачу клинической лабораторной диагностики Клинического центра профилактики и борьбы со СПИД Министерства здравоохранения Краснодарского края; Дмитрию Сергеевичу Колпакову, заведующему клинико-диагностической лабораторией Ростовского НИИ микробиологии и паразитологии; Марине Ридовне Бобковой, заведующей лабораторией вирусов лейкозов НИЦЭМ им. Н.Ф. Гамалеи; Валерии Олеговне Котовой, заведующей лабораторией эпидемиологии и профилактики вирусных гепатитов и СПИДа Хабаровского НИИ эпидемиологии и микробиологии за внесение молекулярной, эпидемиологической и клинической информации в Российскую базу данных устойчивости ВИЧ к антиретровирусным препаратам.

Источник финансирования. Работа выполнена в рамках внутреннего гранта ЦНИИ Эпидемиологии Роспотребнадзора на фундаментальное исследование «Развитие Российской базы данных лекарственной устойчивости ВИЧ-1 к антиретровирусным препаратам».

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Киреев Д.Е., Кириченко А.А., Лопатухин А.Э., Шлыкова А.В., Галкин Н.Ю., Савельер Е.В., Глазов М.Б., Покровский В.В., Акимкин В.Г. Российская база данных устойчивости ВИЧ к антиретровирусным препаратам. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2023;100:online-first.
DOI: https://doi.org/10.36233/0372-9311-345 EDN: https://www.elibrary.ru/rrwanu

Introduction

Molecular and genetic technologies have become indispensable tools in diagnosis, epidemiology, and treatment of infectious diseases. For more than 20 years, methods of identification of nucleotide sequences (NS) of the human immunodeficiency virus (HIV) have been used for detecting drug resistance (DR) of the virus and subsequent assessment of effectiveness of antiretrovirals (ARVs). With the growing number of existing drugs and patients receiving antiretroviral therapy (ART), the significance of the analysis and the number of related studies are increasing. In the meantime, the intermediate results of virus DR studies, namely NS, can also be used in epidemiology to study patterns of the virus spread. In this case, information about the virus can be seen as an additional parameter, which is analyzed together with epidemiological, clinical, and demographic characteristics of a HIV-infected individual. Being an objective parameter, NS are especially useful in studies of diseases difficult to diagnose, chronic, and associated with a high level of stigma.

Therefore, storage of molecular data and related information about a patient, including their further analysis, is an important aspect of the research on epidemiology of HIV infection. There are continuously expanded national and regional HIV NS databases. The most complete national databases in Switzerland [1] and the United Kingdom [2] contain information about more than 50% of HIV-infected residents of these countries. Genetic data on the virus are added to these databases mainly through HIVDR studies performed during routine healthcare delivery to HIV-infected individuals.

The growth rate of the HIV infection epidemic in Russia is much higher than the respective rates in developed countries. In 2021, the reported incidence rate was 48.7 per 100,000 population, while the prevalence rate was 782.0 per 100,000 population. At the end of 2021, there were 1,137,596 people living with HIV in Russia and more than 660,000 people took ARVs. In the meantime, every year, more than 15% of patients taking ARVs do not reach an undetectable viral load[1]. According to the standard healthcare requirement, HIV drug resistance tests should be available to each tenth patient with HIV two times a year[2].

Considering the scale of the epidemic and the requirements of the regulatory documents, Russia needs more than 100,000 HIVDR-related tests annually. However, due to high costs, labor inputs, and insufficient equipment of laboratories at centers for prevention and control of AIDS and infectious diseases (AIDS centers), the actual number of tests is less than 10%. In 2020 and 2021, the total number of reagent kits purchased for HIVDR tests was 7,438 and 7,232, respectively[3]. Therefore, due to the objective reasons, with the same financial support provided to the AIDS centers and the same sequencing technologies used in Russia, the required coverage of HIV-positive individuals by DR tests will not be reached. In addition, the results of the existing few HIV drug resistance tests are often unavailable for further epidemiological studies due to the absence of the tool for collection, storage, and analysis of the information.

To minimize losses of virus-related genetic data and to monitor HIVDR, the Central Research Institute of Epidemiology of Rospotrebnadzor created a Russian database of HIV resistance to ARVs (RDB; https://www.hivresist.ru/). The use of RDB was first governed by the methodological guidelines of 2013[4] and then by the methodological instructions of 2016.[5] Starting from 2021, in accordance with the revised Sanitary Regulations and Standards 3.3686-21[6], all AIDS centers have to deposit NS obtained during HIVDR tests and the related depersonalized information about patients to RDB.

Acquisition of genetic information about the virus and its subsequent bioinformatic analysis can improve the effectiveness of epidemiological surveillance of HIV infection, improve the quality of healthcare delivery and, eventually, contribute to reduction in new cases. The advanced analysis techniques have entailed improvement of RDB.

The **purpose** of the article was to describe the functional capabilities of the Russian HIV antiviral drug resistance database with reference to the NS analysis and related information about patients.

Programming languages and bioinformatic methods

The internal part of RDB is accessible only to registered users and is supported by PHP, JS, HTML, CSS languages. The external part is accessible to any

1 HIV infection in the Russian Federation as of 31/12/2021. AIDS Prevention and Control Report of the Central Research Institute of Epidemiology, Rospotrebnadzor. URL: http://www.hivrussia.info/wp-content/uploads/2022/03/Spravka-VICH-v-Rossii-na-31.12.2021-g.pdf

2 Decree No. 438n of the Ministry of Health of the Russian Federation, 23/6/2022, On Approval of the Standard for the Primary Healthcare for Adults with HIV Infection (Diagnostics, Treatment, and Follow-Up Monitoring).

3 The International Treatment Preparedness Coalition, Eastern Europe, and Central Asia. Analysis of Purchases of Diagnostic Tools for HIV Treatment in Russia in 2020–2021. 2022. URL: https://itpc-eeca.org/wp-content/uploads/2022/07/monitoring-testov-vich-2020-21-gg-1.pdf

4 Methodological guidelines "Monitoring of the Spread of HIV Strains Resistant to Antiviral Drugs". Moscow, 2013.

5 Epidemiological Surveillance of HIV infection. Methodological instructions. Moscow, 2016.

6 Resolution No. 4 of the Chief Public Health Officer of the Russian Federation, 28/1/2021, On Approval of Sanitary Regulations and Standards 3.3686-21 "Sanitary and Epidemiological Requirements for Prevention of Infectious Diseases" (as amended on May 25, 2022)".

DOI: https://doi.org/10.36233/0372-9311-345

user of the Internet and operates on the Bitrix platform.

The statistics is updated automatically at the specified time intervals. The data for the home page are acquired through API (Application Programming Interface) in the json format. Charts displayed on the home page of the website are built using the ChartJS library.

The analysis of NS for identification of subtypes, for the presence of resistance mutations and DR to ARVs was performed using resources of the database of the Stanford University (https://hivdb.stanford.edu/).

The assessment of NS encoding the V3 loop for viral tropism was performed using the resource of the Max Planck Institute for Informatics (https://coreceptor.geno2pheno.org/).

The uploaded NS were screened for contamination using the BLAST (Basic Local Alignment Search) tool. The threshold was set at the 98% and 99% levels of absolute and relative genetic similarity between NS.

Entry of information into the database

There are two options for uploading information. In the first option, information is entered manually for each patient. This option is recommended when data are added to the database regularly.

Different methods of filling fields in a patient card are offered to make information entry easier. The first method includes a drop-down list and a selection of the offered options, thus eliminating the risk of errors associated with manual entry. It is backed up by directories containing information about cities/towns/localities, ARVs, codes related to causes of HIV infection and HIV testing. The second method is used for automatic selection of the region of residence and federal district when the city of residence is entered. The third method is used for selection of multiple ARVs. All methods have prompts and filters. The filters offer the selection between the full and short name of an ARV from the drop-down list.

Any changeable variables such as HIV RNA levels or CD4 cell counts can be added to the patient card at any time. The accuracy of records is constantly monitored; for example, the initial treatment date cannot be later than the date of its completion. All the detected errors are highlighted with red color to be noticeable to the user. Autofill and flash fill features are provided. If any changes in the existing lists are required, including addition of new ARVs, the directories can be amended by RDB administrators.

The second option is used when data on multiple patients must be entered to a spreadsheet that is uploaded to the database. The spreadsheet template can be downloaded by any user and is available in the user account in RDB.

At any time, users can go back to the already added patients to update the information or add new data.

Quality control of information entered into the database

Quality control of epidemiological and clinical information about patients in uploaded spreadsheets

When using spreadsheets for bulk upload, the database provides the following screening features:

- screening for accuracy of filling the template;
- screening for identical reference numbers of patient cards;
- matching the reference numbers of new patient cards against the numbers existing in RDB;
- verification of the new information about the patient against the existing data.

In all cases, if any duplicate is detected, when uploading records of the patient, whose data exist in RDB, the system asks the user to confirm that the new information does not contain any errors, and after the confirmation is received, the new information is added to the existing records. The system searches for duplicate entries by matching reference numbers of cards and by the matching combination of the patient's date of birth, the date of their first positive immunoblot, and the patient's sex.

Quality control of nucleotide sequences

To evaluate the quality of new NS, we have developed an algorithm with the following parameters: the number of the first amino acid coded by NS; the number of the last amino acid coded by NS; the number of reading frame shifts, insertions/deletions, stop codons, degenerate positions, APOBEC, and atypical mutations. New NS are analyzed using the offered algorithm and, if any NS of inadequate quality is detected, are marked by respective flags. The user is offered several options: To delete the flagged NS from the pool intended for upload or to edit and upload the edited NS or to upload NS as they are, without any changes. To prevent any errors in the results of subsequent analyses when they fail the quality control, NS are marked as controversial and are not used in further data analysis and generation of standardized reports.

Contamination screening

To detect any contamination of specimens, which may have occurred during the test, we have developed a program to measure the genetic similarity of the new NS uploaded to RDB. After the fasta file with the new and existing NS is converted into a binary format, genetic similarity is measured for different NS groups (within one group uploaded to the database at the same time; within the group uploaded to the database in the last 3 months; within all NS uploaded by a specific institution). If any abnormally high genetic similarity between the analyzed NS has been identified, the program

marks these NS and notifies the user about possible contamination.

Analysis of entered information, export of data and generation of reports

Analysis of nucleotide sequences

The automated NS analysis is performed using specially developed tools. At present, functional capabilities RDB are implemented through the most practically important set of tools for genetic data analysis. The analysis is performed when HIV NS are uploaded to RDB.

Genetic variants are identified using the algorithm of the Stanford University. RDB exports NS to the university website and then retrieves the information about the specified genetic variants.

NS are assessed for presence of ARV resistance mutations using the Stanford database. The sequence is exported for the remote analysis and then the results of analysis are imported, including automatic translation into Russian language. Upon the completion of the analysis, the HIVDR information is stored in the patient's individual card and can be exported as an Excel spreadsheet or *.pdf file.

Viral tropism assessment is performed in RDB when NS of the HIV gp120 envelope protein V3 loop region are uploaded. The assessment is performed using the geno2pheno algorithm developed at the Max Planck Institute for Informatics [3]. Then the information about viral tropism is added to the patient's individual card.

Data download and report generation

Download functions are available both for all the existing data and for part of the uploaded information. Filters with main parameters can be used for downloaded information, including the date of upload, the date of blood collection, the date of diagnosis, the history of ARV taking, etc. In addition, patient cards can be downloaded in the *.pdf format.

The additional feature includes generation of a standardized report based on the analysis of uploaded data. The system collects the required data in RDB, analyzes these data and generates a standardized annual report. The report contains text fields, which users can edit and update when required; it also has tables containing data on patients and on prevalence of HIVDR mutations in the region. The report also includes statistical data on the sex of patients, the infection routes, HIVDR. Generated reports can be stored for further use or downloaded in the *.pdf format. In addition, users can download statistical graphs to use them in their reports or presentations.

Number of users and amount of uploaded data

On 1/7/2022, a total of 10 users were registered in RDB. The total number of unique records about patients was 10,626 and the total number of NS was 13,126. The numbers of records about patients and NS uploaded by registered users are shown in **Table 1**.

Since molecular and epidemiological studies require clinical and epidemiological information, special attention was given to acquisition of data on the patient's sex, age, region of residence, presumed route of

Table 1. Information about users of the Russian HIV-1 antiviral drug resistance database and the amount of uploaded information

Institution	Number of patients (records)	Number of sequences (genome region)				
		pro-rev	int	pro-rev-int	env	full
Siberian Federal District Center for the Prevention and Control of AIDS, Omsk Research Institute of Natural Focal Infections	1548	1548	0	0	0	0
Nizhny Novgorod Research Institute of Epidemiology and Microbiology named after academician I.N. Blokhina	82	82	0	0	0	0
Yekaterinburg Research Institute of Viral Infections	265	275	0	0	0	0
Krasnoyarsk Regional Center for Prevention and Control of AIDS	62	62	0	0	0	0
Lipetsk Regional Center for the Prevention and Control of AIDS and Infectious Diseases	57	57	0	0	0	0
Clinical Center for the Prevention and Control of AIDS of the Ministry of Health of the Krasnodar Territory	10	10	0	0	0	0
Rostov Research Institute of Microbiology and Parasitology	175	175	0	0	0	0
N.F. Gamaleya Research Center of Epidemiology and Microbiology	3544	3544	24	0	93	0
Central Research Institute for Epidemiology	4801	4729	899	293	857	396
Khabarovsk Research Institute of Epidemiology and Microbiology	82	82	0	0	0	0
Total	10 626	10 564	923	293	950	396

Note. Pro-rev — protease gene NS and 2/3 reverse transcriptase gene NS; int — integrase gene NS; pro-rev-int — protease, reverse transcriptase and integrase gene NS; env — gp120 V3 loop protein NS; full — NS of the entire coding region of HIV genome.

DOI: https://doi.org/10.36233/0372-9311-345

Table 2. Completeness of clinical and epidemiological information in patient records available in RDB

Patient profile		Patient records	
		n	%
Sex	available	10 490	98,7
	no data	136	1,3
Year of birth	available	10 281	96,8
	no data	345	3,2
Year of the first positive immunoblot	available	10 344	97,3
	no data	282	2,7
Alleged route of infection	available	7998	75,3
	no data	2628	24,7
Information about ARV treatment history	available	10 295	96,9
	no data	331	3,1
Region of residence	available	10 143	95,5
	no data	483	4,5
Information about ART regimens	available	4342	81,2
	no data	1005	18,8

infection, and the date of diagnosis as well as the patient's history of taking ARVs and administered ART regimens. Information about such data is shown in **Table 2**.

Copyright and data exchange between users

Any institution that has its user account has an access to all the above mentioned tools of the database and the information uploaded by employees of the institution. The information uploaded by employees of other institutions can also be accessible, provided that the institution has granted access. Thus, on the one hand, RBD ensures confidentiality of the uploaded information. On the other hand, it offers the possibility of fast exchange and joint analysis of collective data when several institutions perform joint scientific and clinical studies.

Conclusion

Collection and storage of genetic information about the pathogen and the related information about patients are of exceptional importance, and at the present stage, the centralized analysis of these data is instrumental in addressing epidemiological issues. For example, the Virus Genome Aggregator of Russia (VGARus) [4], which was created during the COVID-19 pandemic, has demonstrated its indispensability in monitoring of emerging viral variants and in the assessment of their contagiousness.

For HIV infection, there are respective databases, which are used for different purposes. Some researchers use them to analyze general trends of epidemic development [5, 6]; others use them for decisions on epidemic control measures [7]. The most popular HIV infection research areas in Russia, where the database can be of high benefit, include DR studies [8] and the analysis of spread patterns of different genetic variants of the virus [9].

The most important characteristics of any database are the amount of information, which it contains, and functions addressing the specific tasks. They are especially important for performance of molecular and epidemiological studies, as the absence of the required analysis tools turns the database into an ordinary repository, making it similar to spreadsheets, while the small amount of information significantly affects the reliability of research results.

The requirement for uploading HIV NS and the related information about patients to the Russian HIV antiviral drug resistance database in compliance with the Sanitary Regulations and Standards 3.3686-21 offers promising prospects that the amount of uploaded information will increase in future. For this reason, one of the priority tasks is to improve functional capabilities of RDB. Due to the modifications that facilitated the speed and easiness of uploading information, the number of records entered into RDB has significantly increased in the last years. By 1/11/2022, the number of the most important protease and reverse transcriptase gene sequences had increased to more than 10,000. They outnumber the publicly available Russian HIV sequences deposited to the Global HIV Database of the Los Alamos National Laboratory (https://hiv.lanl.gov). Currently, it contains a total of 8,875 sequences deposited from Russia. It should be noted that in RDB, each entry is provided with a much larger amount of epidemiological and clinical information. In the Global HIV Database, the information about some important details such as the date of diagnosis, the region of

residence and ART regimens is absent, while in other databases, the information about the presumed route of transmission (54.2%), the patient's sex (43.9%), and age (16.9%) is missing in many records.

The user-friendly method offered in RDB for entering patients' characteristics helped increase the amount of collected information; the data on the sex, age, date of diagnosis, presumed route of infection, region of residence, and history of ARV taking are filled at 93.4%. Thus, according to the Rospotrebnadzor data on HIV infection in Russia as of 31/12/2021, the percentage of HIV-infected Russian citizens with known HIV NS is 0.66% of the total number of detected cases of infection (1,562,570 cases) and 0.90% of the number of individuals living with HIV (1,137,596 people).

In addition to the increased number of records in RDB, the quality of uploaded information has improved significantly. The automated quality control decreases the risk of errors in clinical and epidemiological characteristics of patients; it also minimizes the risk of uploading false data regarding NS. We have found that up to 5% of the uploaded NS are identical to each other, demonstrating the discrepancy with the epidemiological data of patients. This situation can take place either due to errors during copying and entering data about patients or due to contamination. Here, contamination is defined as contamination of one clinical specimen by another specimen during the test at a laboratory. Regardless of the reason for present of identical HIV NS in different patients, such error can have a disastrous effect. Patients will receive the results of the DR analysis for the virus they are not infected with. The newly developed contamination screening tool eliminates any risk of such errors [10].

Other important, newly added features include automated subtyping, detection of DR mutations and identification of viral tropism. For example, DR data can be used not only at the individual, but also at the population level. Information about the DR structure and prevalence can be used in choosing ART regimens based on national recommendations or in decisions on purchasing of ARVs.

Our further plans include development of a more complex and advanced tool for bioinformatic analysis. Detection of molecular clusters is essential for studies in molecular epidemiology [11, 12]. This function will be available to RDB users in the near future.

Upgrading and updating of RDB will help improve the security of virus-related genetic information obtained from HIVDR tests in Russia and increase the quality of studies in molecular epidemiology of HIV infection. The increased number of tests and continuous expansion of RDB through adding new information will expedite its use in the routine epidemiological practice. The detected molecular clusters will help healthcare specialists identify vulnerable groups and administrative districts characterized by rapid virus transmission, thus contributing to enhancing the effectiveness of epidemic control measures. Thus, in the foreseeable future, RDB can change its status of a solely scientific tool and become a significant component of the epidemiological surveillance.

REFERENCES

1. Scherrer A.U., Traytel A., Braun D.L., et al. Cohort profile update: The Swiss HIV Cohort Study (SHCS). *Int. J. Epidemiol*. 2022;51(1):33–4j. DOI: https://doi.org/10.1093/ije/dyab141
2. Dunn D., Pillay D. UK HIV drug resistance database: background and recent outputs. *J. HIV Ther*. 2007;12(4):97–8.
3. Lengauer T., Sander O., Sierra S. et al. Bioinformatics prediction of HIV coreceptor usage. *Nat. Biotechnol*. 2007;25(12):1407–10. DOI: https://doi.org/10.1038/nbt1371
4. Akimkin V.G., Semenenko T.A., Ugleva S.V., et al. COVID-19 in Russia: epidemiology and molecular genetic monitoring. *Vestnik Rossiyskoy akademii meditsinskikh nauk*. 2022;77(4):254–60. DOI: https://doi.org/10.15690/vramn2121
5. Kusejko K., Salazar-Vizcaya L., Shah C., et al. Sustained effect on hepatitis c elimination among men who have sex with men in the Swiss HIV cohort study: a systematic re-screening for hepatitis C RNA two years following a nation-wide elimination program. *Clin. Infect. Dis*. 2022;75(10):1723–31. DOI: https://doi.org/10.1093/cid/ciac273
6. Kusejko K., Marzel A., Nguyen H., et al. Differences in social and mental well-being of long-term survivors among people who inject drugs and other participants in the Swiss HIV Cohort Study: 1980-2018. *Antivir. Ther*. 2020;25(1):43–54. DOI: https://doi.org/10.3851/IMP3347
7. Poon A.F., Gustafson R., Daly P., et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV*. 2016;3(5):e231–8. DOI: https://doi.org/10.1016/S2352-3018(16)00046-1
8. Ebbert M.T., Mallory M.A., Wilson A.R., et al. Application of a new informatics tool for contamination screening in the HIV sequencing laboratory. *J. Clin. Virol*. 2013;57(3):249–53. DOI: https://doi.org/10.1016/j.jcv.2013.03.013
9. Hassan A.S., Pybus O.G., Sanders E.J., Albert J., Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS*. 2017;31(9):1211–22. DOI: https://doi.org/10.1097/QAD.0000000000001470
10. Kosakovsky Pond S.L., Weaver S., et al. HIV-TRACE (TRAnsmission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens. *Mol. Biol. Evol*. 2018;35(7):1812–9. DOI: https://doi.org/10.1093/molbev/msy016

Информация об авторах

Киреев Дмитрий Евгеньевич✉ — к.б.н., зав. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, dmitkireev@yandex.ru, https://orcid.org/0000-0002-7896-2379

Кириченко Алина Алексеевна — н.с. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-7116-0138

Лопатухин Алексей Эдуардович — н.с. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-2826-699X

Шлыкова Анастасия Вениаминовна — н.с. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-1390-8021

Галкин Николай Юрьевич — старший программист информационных систем центра по развитию информационных технологий и систем ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-2144-8774

Савельер Евгений Викторович — руководитель группы разработки информационных систем центра по развитию информационных технологий и систем ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0003-2811-8097

Глазов Максим Борисович — руководитель центра по развитию информационных технологий и систем ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-2195-1580

Покровский Вадим Валентинович — д.м.н., профессор, академик РАН, зав. специализированным отделом по эпидемиологии и профилактике СПИДа ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0002-9514-7288

Акимкин Василий Геннадьевич — д.м.н., профессор, академик РАН, директор ЦНИИ Эпидемиологии, Москва, Россия, https://orcid.org/0000-0003-4228-9044

Участие авторов. Все авторы внесли существенный вклад в проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию до публикации.

Статья поступила в редакцию 24.11.2022;
принята к публикации 01.02.2023;
опубликована 28.03.2023

Information about the authors

Dmitry E. Kireev✉ — Cand. Sci. (Biol.), Head, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, dmitkireev@yandex.ru, https://orcid.org/0000-0002-7896-2379

Alina A. Kirichenko — researcher, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-7116-0138

Alexey E. Lopatukhin — researcher, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-2826-699X

Anastasia V. Shlykova — researcher, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-1390-8021.

Nikolai Yu. Galkin — senior programmer, Center for the development of information technologies and systems, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-2144-8774

Evgenii V. Saveler — Group Head, Center for the development of information technologies and systems, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0003-2811-8097

Maxim B. Glazov — Head, Center for the development of information technologies and systems, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-2195-1580

Vadim V. Pokrovsky — D. Sci. (Med.), Prof., Full Member of RAS, Head, HIV department, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0002-9514-7288

Vasily G. Akimkin — D. Sci. (Med.), Prof., Full Member of RAS, Director, Central Research Institute for Epidemiology, Moscow, Russia, https://orcid.org/0000-0003-4228-9044

Author contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published.

The article was submitted 24.11.2022;
accepted for publication 01.02.2023;
published 28.03.2023